Decay of correlations for non Hölderian dynamics. A coupling approach*

Xavier Bressaud[†] Roberto Fernández[‡§] Antonio Galves[¶]

Draft

August 23, 2005

Abstract

We present an upper bound on the mixing rate of the equilibrium state of a dynamical systems defined by the one-sided shift and a non Hölder potential of summable variations. The bound follows from an estimation of the relaxation speed of chains with complete connections with summable decay, which is obtained via a explicit coupling between pairs of chains with different histories.

1 Introduction

Let μ_ϕ be the equilibrium state associated to the continuous function ϕ. In this paper we obtain upper bounds for the speed of convergence of the limit

$$\int_X f \circ T^n \, g \, d\mu_\phi \underset{n\to\infty}{\longrightarrow} \int_X f \, d\mu_\phi \int_X g \, d\mu_\phi \tag{1.1}$$

for ϕ with summable variations and T the one-sided shift. We show that this speed is (at least) summable, polynomial or exponential according to the decay rate of the variations of ϕ. The bounds apply for $f \in L^1(\mu_\phi)$ and g with variations decreasing proportionally to those of ϕ.

Previous approaches to the study of the mixing properties of the one-sided shift rely on the use of the transfer operator L_ϕ, defined by the duality,

$$\int_X f \circ T^n \, g \, d\mu_\phi = \int_X f \, L_\phi^n g \, d\mu_\phi \; . \tag{1.2}$$

If ϕ is Hölder, this operator, acting on the subspace of Hölder observables, has a spectral gap and the limit (1.1) is attained at exponential speed (Bowen, 1975). When ϕ is not Hölder, the spectral gap of the transfer operator may vanish and the spectral study becomes rather complicated. To estimate the mixing rate, Kondah, Maume and Schmitt (1996) proved first that the operator is contracting in the Birkhoff projective metric, while Pollicott (1997), following Liverani (1995), considered the transfer operator composed with conditional expectations. In contrast, our approach is based on a probabilistic interpretation of the duality (1.2) in terms of expectations, conditioned with respect to the past, of a chain with complete connections The convergence (1.1) is therefore related to the relaxation properties of this chain. In this paper, such relaxation is studied via a coupling method.

*Work done within the Projeto Temático “Fenômenos Críticos em Processos Evolutivos e Sistemas em Equilíbrio”, supported by FAPESP (grant 95/0790-1), and is part of the activities of the Núcleo de Excelência “Fenômenos Críticos em Probabilidade e Processos Estocásticos” (grant 41.96.0923.00)

[†]Work partially supported by FAPESP (grant 96/04860-7).

[‡]Researcher of the National Research Council (CONICET), Argentina.

[§]Work partially supported by CNPq (grant 301625/95-6).

[¶]Work partially supported by CNPq (grant 301301/79).

Chains with complete connections are processes characterized by having transition probabilities that depend on the whole past in a continuous manner. They were first introduced by Onicescu and Mihoc (1935, 1935a) and soon taken up by Doeblin (1937). These authors proved the first existence and convergence results for these processes, later extended by Harris (1955). [The definition adopted in these works is written in a slightly different manner than current usage in the random-processes literature (see eg. Lalley, 1986). We adopt the latter.] Moreover, their studies were geared towards more complicated objects —called *random systems with complete connections*— where the chain acts as an underlying "index sequence" used to define very general Markov processes. In this form, the chains have ben applied to studies of urn schemes (Onicescu and Mihoc, 1935a), continued fractions (Doeblin, 1940; Iosifescu 1978), learning processes (Karlin, 1953; Iosifescu and Theodorescu, 1969; Norman, 1972) and image coding (Barnsley et al, 1988). As a general reference on the subject we mention the book by Iosifescu and Grigorescu (1990) as well as the historical review presented in Kaijser (1981) and the brief and clear update by Kaijser (1994). These last two references were our main sources for the preceding account. (We thank an anonymous referee for pointing to us key bibliographical sources on the theory of random systems with complete connections.) Our work introduces a novel application of this useful objects to the field of dynamical systems, where they appear in a rather natural way.

Coupling ideas were first introduced by Doeblin in his 1938 work on the convergence to equilibrium of Markov chains. He let two independent trajectories evolve simultaneously, one starting from the stationary measure and the other from an arbitrary distribution. The convergence follows from the fact that both realizations meet at a finite time. [Doeblin published his results in a hardly known paper in the Revue Mathématique de l'Union Interbalkanique. For a description of Doeblin's contributions to probability theory we refer the reader to Lindvall (1991)]. This idea was soon applied to study the existence and relaxation properties of chains with complete connections (Doeblin and Fortet, 1937; Harris, 1955); see also the account by Iosifescu, 1992). Instead of letting the trajectories evolve independently (*trivial* or *product* coupling), one can couple them from the beginning so that they continue to evolve together once they meet. This reduces the "meeting time" and, hence, yields a better rate of convergence. This procedure is what nowadays is known as *coupling* in the stochastic-processes literature. In this setting, it is particularly efficient to use couplings that "load" the diagonal as much as possible. In our work, we apply a particular coupling with this property, sometimes called the *Vaserstein* coupling (eg. in Kaijser, 1981, Lindval, in his 1992 lectures, calls it γ-coupling). For instance, when applied to a Markov processes, this coupling prescription leads to the so-called Dobrushin's ergodic coefficient. The sharpness of the convergence rates provided by different types of Markovian couplings has been recently discussed by Burdzy and Kendall (1998).

The Vasserstein coupling has been applied before to chains with complete connections —for instance by Kaijser (1981, 1994) and by us (Bressaud, Fernández, Galves, 1997)— for different purposes. In this paper we use it to obtain a more careful estimation of the relaxation properties of such chains. As the transition probabilities of the chains considered here depend on the whole past, the coupling can not ensure that two different trajectories will remain equal after their first meeting time. But the coupling has the property that if the trajectories meet they have a large probability of remaining equal, and this probability increases with the number of consecutive agreements. In the summable case, the coupling is such that with probability one the trajectories disagree only a finite number of times. In fact, the approach can also be applied under an assumption weaker than summability [(4.7) below]. This assumption, which was previously put forward by Harris (1955), ensures that trajectories that differ infinitely often do so with a probability of disagreement that goes to zero. The method leads, in particular, to a criterium of uniqueness for g-measures proven by Berbee (1987). The mean time between succesive disagreements provides a bound on the speed of relaxation of the chain and hence, through our probabilistic interpretation of (1.2), of the mixing rate.

Let us mention, as related developments in the context of dynamical systems, the recent papers by Coelho and Collet (1995) and Young (1997). These papers consider the time two independent systems take to become close. This is reminiscent of the coupling ideas.

The paper is organized as follows. The main results and definitions relevant to dynamical systems are stated in Section 2. The relation between chains with complete connections and the transfer operator is spelled out in Section 3. In Section 4, we state and prove the central result on relaxation speeds of chains with complete connections. Theorem 1 on mixing rates for normalized functions is proven in Section 5, while Theorem 2 on rates for the general case is proven in Section 6. The upper bounds on the decay of correlations depend crucially on estimations of the probability of return to the origin of an auxiliary Markov chain, which are presented in Section A.

2 Definitions and statement of the results

Let A be a finite set henceforth called *alphabet*. Let us denote

$$\underline{A} \;=\; \bigl\{x = (x_j)_{j\leq -1}\;,\; x \in A\bigr\} \tag{2.1}$$

the set of sequences of elements of the alphabet indexed by the strictly negative integers. Each sequence $x \in \underline{A}$ will be called a *history*. Given two histories x and y, the notation $x \stackrel{m}{=} y$ indicates that $x_j = y_j$ for all $-m \leq j \leq -1$.

As usual, we endow the set $\underline{A}$ with the product topology and the σ-algebra generated by the cylinder sets. We denote by $\mathcal{C}^0(\underline{A}, \mathbb{R})$ the space of real-valued continuous functions on $\underline{A}$.

We consider the one-sided shift T on $\underline{A}$,

$$\begin{array}{rccl} T: & \underline{A} & \longrightarrow & \underline{A} \\ & x & \longmapsto & T(x) = (x_{i-1})_{i\leq -1}. \end{array}$$

Given an element a in A and an element x in $\underline{A}$, we shall denote by xa the element z in $\underline{A}$ such that $z_{-1} = a$ and $T(z) = x$.

Given a function ϕ on $\underline{A}$, $\phi : \underline{A} \to \mathbb{R}$, we define its sequence of *variations* $(\mathrm{var}_m(\phi))_{m\in\mathbb{N}}$,

$$\mathrm{var}_m(\phi) = \sup_{x \stackrel{m}{=} y} |\phi(x) - \phi(y)|\ . \tag{2.2}$$

We shall say that it has *summable variations* if,

$$\sum_{m\geq 1} \mathrm{var}_m(\phi) < +\infty\ , \tag{2.3}$$

and that it is *normalized* if it satisfies,

$$\forall x \in \underline{A}\ ,\quad \sum_{a\in A} e^{\phi(xa)} = 1\ . \tag{2.4}$$

We say that a shift-invariant measure μ on $\underline{A}$ is *compatible with the normalized function* ϕ if and only if, for μ_ϕ-almost-all x in $\underline{A}$,

$$\mathbf{E}_{\mu_\phi}\bigl(1_{\{x_{-1}=a\}}|\mathcal{F}_{\leq -2}\bigr)(x) \;=\; e^{\phi(T(x)a)}\ , \tag{2.5}$$

where the left-hand side is the usual conditional expectation of the the indicator function of the event $\{x_{-1} = a\}$ with respect to the σ-algebra of the past up to time -2.

An equivalent way of expressing this is by saying that μ_ϕ is a g-measure for $g = e^\phi$. If ϕ has summable variations, and even under a slightly weaker conditions, then such a measure is unique and will be denoted μ_ϕ. The measure μ_ϕ can also be characterized via a variational principle, in which context it is called *equilibrium state* for ϕ. For details see Ledrappier (1974), Walters (1975), Quas (1996) and Berbee (1987).

For a non-constant ϕ, we consider the seminorm

$$||g||_\phi \;=\; \sup_{k\geq 0} \frac{\mathrm{var}_k(g)}{\mathrm{var}_k(\phi)} \tag{2.6}$$

and the subspace of $\mathcal{C}^0(\underline{A})$ defined by,

$$V_\phi \;=\; \Big\{ g \in \mathcal{C}^0(\underline{A},\mathbb{R})\,,\; ||g||_\phi < +\infty \Big\}\;. \tag{2.7}$$

Given a real-valued sequence $(\gamma_n)_{n\in\mathbb{N}}$, let $(S_n^{(\gamma)})_{n\in\mathbb{N}}$ be the Markov chain taking values in the set $\mathbb{N}$ of natural numbers starting from the origin

$$\mathbf{P}(S_0^{(\gamma)} = 0) \;=\; 1 \tag{2.8}$$

whose transition probabilities are defined by

$$\begin{aligned} p_{i,i+1} &= 1-\gamma_i \\ p_{i,0} &= \gamma_i\;, \end{aligned} \tag{2.9}$$

for all $i \in \mathbb{N}$. For any $n \geq 1$ we define

$$\gamma_n^* = \mathbf{P}(S_n^{(\gamma)} = 0)\;. \tag{2.10}$$

We now state our first result.

Theorem 1 *Let $\phi : \underline{A} \to \mathbb{R}$ be a normalized function with summable variations and set*

$$\gamma_n \;=\; 1 - e^{-\mathrm{var}_n(\phi)}\;. \tag{2.11}$$

Then,

$$\left| \int f\circ T^n\, g\, d\mu_\phi - \int f\, d\mu_\phi \int g\, d\mu_\phi \right| \;\leq\; ||f||_1\, ||g||_\phi \sum_{k=0}^{n} \mathrm{var}_k(\phi)\, \gamma_{n-k}^* \tag{2.12}$$

$$\leq\; C\, ||f||_1\, ||g||_\phi\, \gamma_n^*\;, \tag{2.13}$$

for all $f \in L^1(\mu_\phi)$ and $g \in V_\phi$, for a computable constant C.

This theorem is proven in Section 5, using the results obtained in Section 4 on the relaxation speed of chains with complete connections.

For each non-normalized function ϕ with summable variations there exists a unique positive function ρ such that the function

$$\psi = \phi + \log\rho - \log\rho\circ T \tag{2.14}$$

is normalized (Walters, 1975). We call ψ the *normalization* of ϕ. The construction of compatible measures given in (2.5) looses its meaning for non-normalized ϕ. It is necessary to resort to an alternative characterization in terms of a variational principle (see eg. Bowen 1975) leading to equilibrium states. In Walters (1975) it is proven that:

(a) ϕ with summable variations admits a unique equilibrium state, that we denote also μ_ϕ;

(b) the corresponding normalized ψ, given by (2.14), admits a unique compatible measure μ_ψ (even when the variations of ψ may not be summable), and

(c) $\mu_\phi = \mu_\psi$.

Our second theorem generalizes Theorem 1 to non-normalized functions.

Theorem 2 *Let $\phi : \underline{A} \to \mathbb{R}$ be a function with summable variations and let ψ be its normalization. Let $(n_m)_{m\in\mathbb{N}}$ be an increasing subadditive sequence such that the subsequence of the rests, $\big(\sum_{k\geq n_m} \mathrm{var}_k(\phi)\big)_{m\geq 0}$, is summable, and*

$$\overline{\gamma}_m = 1 - e^{-3\sum_{k\geq n_m} \mathrm{var}_k(\phi)} \; ; \tag{2.15}$$

then,

$$\left|\int f\circ T^n\, g\, d\mu_\phi - \int f\, d\mu_\phi \int g\, d\mu_\phi\right| \;\leq\; ||f||_1\, ||g||_\phi \sum_{k=0}^{n} \mathrm{var}_{n_k}(\phi)\, \overline{\gamma}^*_{n-k} \tag{2.16}$$

$$\leq\; C\, ||f||_1\, ||g||_\phi\, \overline{\gamma}^*_n \; , \tag{2.17}$$

for all $f \in L^1(\mu_\phi)$ and $g \in V_\phi$, for a computable constant C. Here $\overline{\gamma}^*$ is defined as in (2.10) but using the sequence $(\overline{\gamma}^*_n)_{n\in\mathbb{N}}$.

The estimation of the large-n behavior of the sequence $(\gamma^*_n)_{n\in\mathbb{N}}$ given the behavior of the original $(\gamma_n)_{n\in\mathbb{N}}$ only requires elementary computations. For the convenience of the reader we summarize some results in Appendix A.

3 Transfer operators and chains.

Let P be a family of transition probabilities on $A \times \underline{A}$,

$$\begin{array}{rccl} P: & A\times\underline{A} & \longrightarrow & [0;1] \\ & (a,z) & \longmapsto & P(a\,|\,z)\; . \end{array} \tag{3.1}$$

Given a history x, a *chain with past x and transitions P*, is the process $(Z^x_n)_{n\in\mathbf{Z}}$ whose conditional probabilities satisfy

$$\mathbf{P}(Z^x_n = a\,|\,Z^x_{n+j} = z_j, j\leq -1) \;=\; P(a\,|\,z) \text{ for } n\geq 0\;, \tag{3.2}$$

for all $a\in A$ and all histories z with $z_{j-n} = x_j, j\leq -1$, and such that

$$Z^x_n = x_n\;, \quad \text{for } n\leq -1\;. \tag{3.3}$$

This chain can be interpreted as a conditioned version of the process defined by the transition probabilities (3.1), given a past x (for more details, see Quas 1996).

Let $\phi : \underline{A} \to \mathbb{R}$ be a continuous normalized function. The transfer operator associated to ϕ is the operator L_ϕ acting on $\mathcal{C}^0(\underline{A}, \mathbb{R})$ defined by,

$$L_\phi f(x) = \sum_{y\,:\,T(y)=x} e^{\phi(y)} f(y)\; . \tag{3.4}$$

This operator is related to the conditional probability (2.5) in the form

$$\mathbf{E}_{\mu_\phi}\big(f\,|\,\mathcal{F}_{\leq -2}\big) \;=\; \big(L_\phi f\big)\circ T\; . \tag{3.5}$$

This relation shows the equivalence of (1.2) and (3.4) as definitions of the operator. In addition, if ϕ is normalized we can construct, for each history $x\in\underline{A}$, the chain $Z^x_\phi = (Z^x_n)_{n\in\mathbf{Z}}$ with past x and transition probabilities

$$P(a\,|\,x) = e^{\phi(xa)}\; . \tag{3.6}$$

Iterates of the transfer operator, $L_\phi^n g(x)$, on functions $g \in \mathcal{C}^0(\underline{A})$ can be interpreted as expectations $\mathbf{E}[g((Z^x_{n+j})_{j\le -1})]$ of the chain. Indeed,

$$\begin{aligned}
L_\phi^n g(x) &= \sum_{a_1,\ldots,a_n\in A} e^{\sum_{k=1}^n \phi(xa_1\cdots a_k)} g(xa_1\cdots a_n)\\
&= \sum_{a_1,\ldots,a_n\in A} \left(\prod_{k=1}^n P(a_k \,|\, a_{k-1}\cdots a_1 x)\right) g(xa_1\cdots a_n)\\
&= \mathbf{E}[g((Z^x_{n+j})_{j\le -1})]\ .
\end{aligned}$$

¿From this expression and the classical duality (1.2) between the composition by the shift and the transfer operator L_ϕ in $L^2(\mu_\phi)$, we obtain the following expression for the decay of correlations,

$$\begin{aligned}
&\int f\circ T^n\, g\, d\mu_\phi - \int f\, d\mu_\phi \int g\, d\mu_\phi\\
&= \int f(x)\, L_\phi^n g(x)\, d\mu_\phi(x) - \int f(x) \left(\int L_\phi^n g(y)\, d\mu_\phi(y)\right) d\mu_\phi(x)\\
&= \int f(x) \int \Big(\mathbf{E}[g((Z^x_{n+j})_{j\le -1})] - \mathbf{E}[g((Z^y_{n+j})_{j\le -1})]\Big)\, d\mu_\phi(y)\, d\mu_\phi(x)\ . \qquad (3.7)
\end{aligned}$$

This inequality shows how the speed of decay of correlations can be bounded by the speed with which the chain loosses its memory. We deal with the later problem in the next section.

4 Relaxation speed for chains with complete connections

4.1 Definitions and main result

We consider chains whose transition probabilities satisfy

$$\inf_{x,y:x\overset{m}{=}y} \frac{P(a\,|\,x)}{P(a\,|\,y)} \ \ge\ 1-\gamma_m\ , \qquad (4.1)$$

for some real-valued sequence $(\gamma_m)_{m\in\mathbb{N}}$, decreasing to 0 as m tends to $+\infty$. Without loss of generality, this decrease can be assumed to be monotonic. To avoid trivialities we assume $\gamma_0 < 1$. In the literature, a *stationary* process satisfying (4.1) is called a *chain with complete connections.*

For a set of transition probabilities satisfying (4.1), we consider, for each $x \in \underline{A}$, the chain $(Z^x_n)_{n\in\mathbb{Z}}$ with past x and transitions P [see (3.2)–(3.3)]. The following proposition plays a central role in the proof of our results.

Proposition 1 *For all histories $x, y \in \underline{A}$, there is a coupling $\big((\widetilde{U}^{x,y}_n, \widetilde{V}^{x,y}_n)\big)_{n\in\mathbb{Z}}$ of $(Z^x_n)_{n\in\mathbb{Z}}$ and $(Z^y_n)_{n\in\mathbb{Z}}$ such that the integer-valued process $(T^{x,y}_n)_{n\in\mathbf{Z}}$ defined by*

$$T^{x,y}_n \ =\ \inf\{m\ge 0 \,:\, \widetilde{U}^{x,y}_{n-m} \ne \widetilde{V}^{x,y}_{n-m}\}, \qquad (4.2)$$

satisfies

$$\mathbf{P}(T^{x,y}_n = 0) \ \le\ \gamma^*_n \qquad (4.3)$$

for $n\ge 0$, where γ^*_n was defined in (2.10).

The proof of this proposition is given in Section 4.4.

An immediate consequence of this proposition is the following bound on the relaxation rate of the processes Z^x.

Corollary 1 *For all histories x and y, for all $a \in A$,*

$$\Big|\mathbf{P}(Z_n^x = a) - \mathbf{P}(Z_n^y = a)\Big| \;\le\; \gamma_n^* \,, \tag{4.4}$$

and, for $k \ge 1$,

$$\begin{aligned}\Big|\mathbf{P}\Big((Z_n^x,\ldots,Z_{n+k}^x) = (a_0,\ldots,a_k)\Big) - \mathbf{P}\Big((Z_n^y,\ldots,Z_{n+k}^y) = (a_0,\ldots,a_k)\Big)\Big| \\ \le \;\; \sum_{j=0}^{k}\left(\prod_{m=1}^{j-1}(1-\gamma_m)\right)\gamma_{n-j}^* \,. \end{aligned} \tag{4.5}$$

This lemma is proved in Section 4.5.

Remark 1 Whenever

$$\gamma_n^* \to 0 \,, \tag{4.6}$$

inequality (4.4) implies the existence and uniqueness of the invariant measure compatible with a system of conditional probabilities satisfying (4.1). In fact, property (4.6) holds under the condition

$$\sum_{m\ge1}\prod_{k=0}^{m}(1-\gamma_k) \;=\; +\infty \,., \tag{4.7}$$

which is weaker than summability. In this case, the Markov chain $(S_n^{(\gamma)})_{n\in\mathbb{N}}$ is no longer transient but it is null recurrent and the property $\mathbf{P}(S_n^{(\gamma)} = 0) \to 0$ remains true.

Remark 2 If $X = (X_n)_{n\in\mathbf{Z}}$ is a stationary process with transition P satisfying (4.1), then Corollary 1 implies

$$\Big|\mathbf{P}(Z_n^x = a) - \mathbf{P}(X_n = a)\Big| \;\le\; \gamma_n^* \,, \tag{4.8}$$

uniformly in the history x.

4.2 Maximal coupling

Given two probability distributions $\mu = (\mu(a))_{a\in A}$ and $\nu = (\nu(a))_{a\in A}$ we denote by $\mu\tilde{\times}\nu = (\mu\tilde{\times}\nu(a,b))_{(a,b)\in A\times A}$ the so-called *maximal coupling* of the distributions μ and ν defined as follows:

$$\begin{cases} \mu\tilde{\times}\nu(a,a) = \mu(a)\wedge\nu(a) & \text{if } a = b \\ \\ \mu\tilde{\times}\nu(a,b) = \dfrac{(\mu(a)-\nu(a))^+(\nu(b)-\mu(b))^+}{\sum_{e\in A}(\mu(e)-\nu(e))^+} & \text{if } a \ne b\,. \end{cases} \tag{4.9}$$

For more details on maximal couplings see Appendix A.1 in Barbour, Holst and Janson (1992).

The coupling is maximal in the sense that the distribution $\mu\tilde{\times}\nu$ on $A\times A$ maximizes the weight

$$\Delta(\zeta) = \sum_{a\in A}\zeta(a,a)$$

of the diagonal among the distributions ζ on $A\times A$ satisfying simultaneously

$$\sum_{a\in A}\zeta(a,b) = \nu(b) \quad \text{and} \quad \sum_{b\in A}\zeta(a,b) = \mu(a)\,.$$

For this coupling, the weight $\Delta(\mu\tilde{\times}\nu)$ of the diagonal satisfies,

$$\Delta(\mu\tilde{\times}\nu) = \sum_{a\in A}\mu(a)\wedge\nu(a) = 1 - \sum_{a\in A}(\mu(a)-\nu(a))^+ = 1 - \frac{1}{2}\sum_{a\in A}|\mu(a)-\nu(a)|. \tag{4.10}$$

Moreover,

$$\Delta(\mu\tilde{\times}\nu) = 1 - \sum_{a\in A}\mu(a)\left(1-\frac{\nu(a)}{\mu(a)}\right)^+ \ge 1 - \sum_{a\in A}\mu(a)\left(1-\inf_{a'\in A}\frac{\nu(a')}{\mu(a')}\right) = \inf_{a\in A}\frac{\nu(a)}{\mu(a)}. \tag{4.11}$$

4.3 Coupling of chains with different pasts

Given a double history (x, y), we consider the transition probabilities defined by the maximal coupling

$$\widetilde{P}((a,b)\,|\,x,y) = \left[P(\cdot\,|\,x)\tilde{\times}P(\cdot\,|\,y)\right](a,b)\ . \tag{4.12}$$

By (4.1) we have,

$$\inf_{a\in A, u\overset{m}{=}v} \frac{P(a\,|\,u)}{P(a\,|\,v)} \ \geq\ 1-\gamma_m.$$

By (4.11) this implies that

$$\Delta\left(\widetilde{P}(\,\cdot\,,\,\cdot\,|\,x,y)\right) \ \geq\ 1-\gamma_m\ , \tag{4.13}$$

whenever $x \overset{m}{=} y$.

Now, we fix a double history (x, y) and we define $\big((\widetilde{U}_n^{x,y}, \widetilde{V}_n^{x,y})\big)_{n\in\mathbb{Z}}$ to be the chain taking values in A^2, with past (x, y) and transition probabilities given by (4.12). If $x \overset{m}{=} y$, (4.13) yields

$$\mathbf{P}(\widetilde{U}_0^{x,y} \neq \widetilde{V}_0^{x,y}) \ \leq\ \gamma_m. \tag{4.14}$$

We denote

$$\Delta_{m,n} \ :=\ \left\{\widetilde{U}_j = \widetilde{V}_j\ ,\ m \leq j \leq n\right\}\ . \tag{4.15}$$

Notice that $\Delta_{-m,-1}$ is the reunion over all the sequences x, y with $x \overset{m}{=} y$ of the events $\{(\widetilde{U}_j, \widetilde{V}_j) = (x_j, y_j)\,;\, j \leq -1\}$. Using the stationarity of the conditional probabilities, we obtain

$$\mathbf{P}(\widetilde{U}_n \neq \widetilde{V}_n\,|\,\Delta_{n-m,n-1}) \ \leq\ \gamma_m\ , \tag{4.16}$$

for all $n \geq 0$.

4.4 Proof of Proposition 1

¿From this subsection on, will be working with bounds which are uniform in x, y, hence we will omit, with a few exceptions, the superscript x, y in the processes $T_n^{x,y}$ (defined below), $\widetilde{U}_n^{x,y}$ and $\widetilde{V}_n^{x,y}$.

Let us consider the integer-valued process $(T_n)_{n\in\mathbf{Z}}$ defined by:

$$T_n \ =\ \inf\{m \geq 0\ :\ \widetilde{U}_{n-m} \neq \widetilde{V}_{n-m}\}\ . \tag{4.17}$$

For each time n, the random variable T_n counts the number of steps backwards needed to find a difference in the coupling. First, notice that (4.16) implies that,

$$\mathbf{P}(T_{n+1} = k+1\,|\,T_n = k) \ \geq\ 1-\gamma_k \tag{4.18}$$

and

$$\mathbf{P}(T_{n+1} = 0\,|\,T_n = k) \ \leq\ \gamma_k\ , \tag{4.19}$$

all the other transition probabilities being zero. This process $(T_n)_{n\in\mathbf{Z}}$ is not a Markov chain.

We now consider the integer-valued Markov chain $(S_n^{(\gamma)})_{n\geq 0}$ starting from state 0 and with transition probabilities given by (2.9), that is $p_{i,i+1} = 1-\gamma_i$ and $p_{i,0} = \gamma_i$. Proposition 1 follows from the following lemma, setting $k = 1$.

Lemma 1 *For each $k \in \mathbf{N}$, the following inequality holds:*

$$\mathbf{P}(S_n^{(\gamma)} \geq k) \ \leq\ \mathbf{P}(T_n \geq k) \tag{4.20}$$

Proof We shall proceed by induction on n. Since $\mathbf{P}(S_0^{(\gamma)} = 0) = 1$, inequalities (4.20) holds for $n = 0$. Assume now that (4.20) holds for some integer n. There is nothing to prove for $k = 0$.. For $k \geq 1$,

$$\begin{aligned}
\mathbf{P}(T_{n+1} \geq k) &= \sum_{m=k}^{+\infty} \mathbf{P}(T_{n+1} = m) \\
&= \sum_{m=k}^{+\infty} \mathbf{P}(T_{n+1} = m \,|\, T_n = m-1)\, \mathbf{P}(T_n = m-1) \\
&\geq \sum_{m=k}^{+\infty} (1 - \gamma_{m-1})\, \mathbf{P}(T_n = m-1) \\
&= \sum_{m=k}^{+\infty} (1 - \gamma_{m-1}) \Big(\mathbf{P}(T_n \geq m-1) - \mathbf{P}(T_n \geq m) \Big) \\
&= (1 - \gamma_{k-1})\, \mathbf{P}(T_n \geq k-1) + \sum_{m=k}^{+\infty} (\gamma_{m-1} - \gamma_m)\, \mathbf{P}(T_n \geq m) \, . \qquad (4.21)
\end{aligned}$$

By the same computation, we see that

$$\mathbf{P}(S_{n+1}^{(\gamma)} \geq k) \;=\; (1 - \gamma_{k-1})\, \mathbf{P}(S_n^{(\gamma)} \geq k-1) + \sum_{m=k}^{+\infty} (\gamma_{m-1} - \gamma_m)\, \mathbf{P}(S_n^{(\gamma)} \geq m) \, . \qquad (4.22)$$

Hence, using the recurrence assumption and the fact that $(\gamma_n)_{n \geq 0}$ is decreasing we conclude that

$$\mathbf{P}(T_{n+1} \geq k) \;\geq\; \mathbf{P}(S_{n+1}^{(\gamma)} \geq k) \, ,$$

for all $k \geq 1$. $\square$

4.5 Proof of Corollary 1

To prove (4.4), first notice that by construction the process $(\widetilde{U}_n)_{n \in \mathbb{Z}}$ has the same law as $(Z_n^x)_{n \in \mathbb{Z}}$ and $(\widetilde{V}_n)_{n \in \mathbb{Z}}$ has the same law as $(Z_n^y)_{n \in \mathbb{Z}}$. Thus,

$$\Big| \mathbf{P}(Z_n^x = a) - \mathbf{P}(Z_n^y = a) \Big| \;=\; \Big| \mathbf{P}(\widetilde{U}_n = a) - \mathbf{P}(\widetilde{V}_n = a) \Big| \;\leq\; \mathbf{P}(\widetilde{U}_n \neq \widetilde{V}_n)) \qquad (4.23)$$

Hence, by definition of the process T_n and Lemma 1,

$$\Big| \mathbf{P}(Z_n^x = a) - \mathbf{P}(Z_n^y = a) \Big| \;\leq\; \mathbf{P}(T_n = 0) \leq \mathbf{P}(S_n^{(\gamma)} = 0) \, . \qquad (4.24)$$

The proof of (4.5) starts similarly:

$$\begin{aligned}
&\Big| \mathbf{P}\Big((Z_n^x, \ldots, Z_{n+k}^x) = (a_0, \ldots, a_k) \Big) - \mathbf{P}\Big((Z_n^y, \ldots, Z_{n+k}^y) = (a_0, \ldots, a_k) \Big) \Big| \\
&\qquad \leq \; \mathbf{P}(S_{n+k}^{(\gamma)} \leq k+1).
\end{aligned}$$

To conclude, we notice that,

$$\mathbf{P}(S_n^{(\gamma)} \leq k) \;=\; \sum_{j=0}^{k} \mathbf{P}(S_n^{(\gamma)} = j) \;=\; \sum_{j=0}^{k} \left(\prod_{m=1}^{j-1} (1 - \gamma_m) \right) \mathbf{P}(S_{n-j}^{(\gamma)} = 0) \, . \square \qquad (4.25)$$

5 Proof of Theorem 1

The proof of Theorem 1 is based on the inequality

$$\left| \int f \circ T^n g d\mu - \int f d\mu \int g d\mu \right| \;\leq\; ||f||_1 \sup_{x,y} \mathbf{E} \left[\left| g((\tilde{U}_{n+j}^{x,y})_{j \leq -1}) - g((\tilde{V}_{n+j}^{x,y})_{j \leq -1}) \right| \right] , \qquad (5.1)$$

which follows from (3.7) and the fact that $\big((\widetilde{U}^{x,y},\widetilde{V}^{x,y})\big)_{n\in\mathbb{Z}}$ is a coupling between the chains with pasts x and y, respectively. An upper bound to the right-hand side is provided by Proposition 1. We see that the transition probabilities (3.6) satisfy condition (4.1), since

$$\frac{P(a\,|\,x)}{P(a\,|\,y)} \;=\; e^{\phi(ax)-\phi(ay)} \;\geq\; e^{-\mathrm{var}_{m+1}(\phi)} \tag{5.2}$$

whenever $x, y \in \underline{A}$ are such that $x \stackrel{m}{=} y$ for some $m \in \mathbb{N}$. We can therefore apply Proposition 1 with

$$\gamma_m \;=\; 1 - e^{-\mathrm{var}_{m+1}(\phi)}\;, \tag{5.3}$$

which tends monotonically to zero if $\sum_{m\geq 1}\mathrm{var}_m(\phi) < +\infty$.

To prove (2.12) we use the process $(T_n^{x,y})_{n\in\mathbf{Z}}$ to obtain the upper bound

$$\begin{aligned}
\mathbf{E}\Big[\Big|g((\widetilde{U}^{x,y}_{n+j})_{j\leq -1}) - g((\widetilde{V}^{x,y}_{n+j})_{j\leq -1})\Big|\Big] \;&=\; \mathbf{E}\Bigg[\sum_{k=0}^{+\infty} 1_{\{T_n^{x,y}=k\}}\Big|g((\tilde{U}_{n+j})_{j\leq -1}) - g((\tilde{V}_{n+j})_{j\leq -1})\Big|\Bigg]\\
&\leq\; \sum_{k=0}^{+\infty}\mathrm{var}_k(g)\,\mathbf{P}(T_n^{x,y}=k)\\
&\leq\; ||g||_\phi \sum_{k=0}^{+\infty}\mathrm{var}_k(\phi)\,\mathbf{P}(T_n^{x,y}=k)\;.
\end{aligned} \tag{5.4}$$

Now, in order to use the bound (4.3) of Proposition (1) we resort to the monotonicity of the variations of ϕ:

$$\begin{aligned}
\sum_{k=0}^{+\infty}\mathrm{var}_k(\phi)\,\mathbf{P}(T_n^{x,y}=k) \;&\leq\; \sum_{k=0}^{n-1}\mathrm{var}_k(\phi)\,\mathbf{P}(T_n^{x,y}=k) + \mathrm{var}_n(\phi)\sum_{k=n}^{+\infty}\mathbf{P}(T_n^{x,y}=k)\\
&=\; \sum_{k=0}^{n-1}\mathrm{var}_k(\phi)\mathbf{P}(T_{n-k}^{x,y}=0) + \mathrm{var}_n(\phi)\sum_{k=n}^{+\infty}\mathbf{P}(T_0^{x,y}=k-n)\\
&\leq\; \sum_{k=0}^{n}\mathrm{var}_k(\phi)\,\mathbf{P}(S_{n-k}^{(\gamma)}=0)\;,
\end{aligned} \tag{5.5}$$

uniformly in x, y. The bound (2.12) follows from (5.1), (5.4), (5.5) and the fact that

$$\sum_{j=0}^{+\infty}\mathbf{P}(T_0^{x,y}=j) \;=\; 1 \;=\; \mathbf{P}(S_0^{(\gamma)}=0)\;. \tag{5.6}$$

To prove (2.13) we use the strong Markov poroperty of the process $(S_n^{(\gamma)})_{n\in\mathbb{N}}$ to obtain

$$\mathbf{P}(S_n^{(\gamma)}=0) = \sum_{k=1}^{n}\mathbf{P}(\tau=k)\,\mathbf{P}(S_{n-k}^{(\gamma)}=0)\;, \tag{5.7}$$

where

$$\tau \;=\; \inf\{n>0; S_n^{(\gamma)}=0\}\;. \tag{5.8}$$

We now use (5.7) to bound the last line in (5.5) in the form

$$\begin{aligned}
\sum_{k=0}^{n}\mathrm{var}_k(\phi)\,\mathbf{P}(S_{n-k}^{(\gamma)}=0) \;&\leq\; \sum_{k=1}^{n}\big[\mathrm{var}_0(\phi)\,\mathbf{P}(\tau=k)+\mathrm{var}_k(\phi)\big]\,\mathbf{P}(S_{n-k}^{(\gamma)}=0)\\
&\leq\; C\sum_{k=1}^{n}\mathbf{P}(\tau=k)\,\mathbf{P}(S_{n-k}^{(\gamma)}=0)\\
&=\; C\,\mathbf{P}(S_n^{(\gamma)}=0)\;,
\end{aligned} \tag{5.9}$$

with

$$C \;=\; \mathrm{var}_0(\phi) + \sup_k \frac{\mathrm{var}_k(\phi)}{\mathbf{P}(\tau = k)}\;. \tag{5.10}$$

To conclude, we must prove that the constant C is finite. By direct computation,

$$\begin{aligned}
\mathbf{P}(\tau = 1) \;&=\; \gamma_0, \\
\mathbf{P}(\tau = n) \;&=\; \gamma_{n-1} \prod_{m=0}^{n-2} (1-\gamma_m) \quad \text{ for } n \geq 2, \\
\mathbf{P}(\tau = +\infty) \;&=\; \prod_{m=0}^{+\infty} (1-\gamma_m)\;.
\end{aligned} \tag{5.11}$$

¿From this and (2.11) we obtain

$$\lim_{k\to\infty} \frac{\mathrm{var}_k(\phi)}{\mathbf{P}(\tau = k)} \;=\; \lim_{k\to\infty} \frac{\mathrm{var}_k(\phi)}{1-e^{-\mathrm{var}_k(\phi)}} \, \frac{1}{\prod_{m=0}^{k-2}(1-\gamma_m)}\;. \tag{5.12}$$

Since $\mathrm{var}_k(\phi) \to 0$, the first fraction converges to 1. We see from (5.11) that the second fraction converges to $1/\mathbf{P}(\tau = +\infty)$. By elementary calculus, this is finite since ϕ has summable variations. □

Remark 3 The previous computations lead to stronger results for more regular functions g. For example, when g satisfies

$$\mathrm{var}_k(g) \;\leq\; ||g||_\theta \, \theta^k \tag{5.13}$$

for some $\theta < 1$ and some $||g||_\theta < \infty$ (Hölder norm of g), a chain of inequalities almost identical to those ending in (5.4) leads to

$$\begin{aligned}
\left| \int f \circ T^n g d\mu - \int f d\mu \int g d\mu \right| \;&\leq\; ||f||_1 \sum_{k=0}^{+\infty} ||g||_\theta \, \theta^k \, \gamma^*_{n-k} \\
&\leq\; ||f||_1 \, ||g||_\theta \, \theta^n \sum_{k=0}^{n} \theta^{-k} \, \gamma^*_k\;.
\end{aligned} \tag{5.14}$$

On the other hand, if g is a function that depends only on the first coordinate, we get,

$$\begin{aligned}
\left| \int f \circ T^n g d\mu - \int f d\mu \int g d\mu \right| \;&\leq\; ||f||_1 \sup_{x,y} \left| \mathbf{E}[g(Z_n^x)] - \mathbf{E}[g(Z_n^y)] \right| \\
&\leq\; ||f||_1 \, ||g||_\infty \mathbf{P}(\tilde{U}_n \neq \tilde{V}_n) \\
&\leq\; ||f||_1 \, ||g||_\infty \, \gamma^*_n\;.
\end{aligned} \tag{5.15}$$

6 Proof of Theorem 2

We now consider the general case where the function ϕ is not necessarily normalized. In this case we resort to the normalization ψ define in (2.14) and we consider chains with transition probabilities

$$P(a \,|\, x) \;=\; e^{\phi(xa)} \, \frac{\rho(xa)}{\rho(x)} \;=:\; e^{\psi(xa)}\;. \tag{6.1}$$

However, the summability of the variations of ϕ does not imply the analogous condition for ψ, because there are addition "oscillations" due to the cocycle $\log \rho - \log \rho \circ T$. Instead,

$$\left. \begin{array}{r} \mathrm{var}_m \psi \\ \mathrm{var}_m(\log \rho) \end{array} \right\} \;\leq\; \sum_{k \geq m} \mathrm{var}_k(\phi)\;, \tag{6.2}$$

for all $m \geq 0$ (see Walters 1978). Hence, we can apply Theorem 1 only under the condition

$$\sum_{k=1}^{+\infty} k \, \mathrm{var}_k(\phi) < +\infty\;. \tag{6.3}$$

If this is the case, the correlations for functions $f \in L^1(\mu)$ and $g \in V_\psi$ decay faster than γ^*_m, where $\gamma_m = e^{\sum_{k\geq m} \mathrm{var}_k(\phi)} - 1$.

To prove the general result without assuming (6.3) we must work with *block* transition probabilities, which are less sensitive to the oscillations of the cocycle. More precisely, given a family of transition probabilities P on $A \times \underline{A}$, let P_n denote the corresponding transition probabilities on $A^n \times \underline{A}$:

$$P_{n+1}(a_{0,n} \,|\, x) \;=\; P(a_n \,|\, a_{n-1} \cdots a_1 x) \cdots P(a_2 \,|\, a_1 x)\, P(a_1 \,|\, x) \tag{6.4}$$

where

$$a_{0,n} \;:=\; (a_0, \ldots, a_n) \;\in A^{n+1} \;. \tag{6.5}$$

If the transition probabilities P are defined by a normalized function ϕ as in (3.6), then we see from (6.4) that the transition probabilities P_n obey a similar relation

$$P_n(a_{0,n-1} \,|\, x) = e^{\phi_n(x a_{0,n-1})} \;, \tag{6.6}$$

with

$$\phi_n(x a_{0,n-1}) \;:=\; \sum_{k=0}^{n-1} \phi(x a_0 \cdots a_k) \;. \tag{6.7}$$

In particular, for transitions (6.1) the formula (6.4) yields

$$\psi_n \;=\; \phi_n + \log\rho - \log\rho \circ T^n \;. \tag{6.8}$$

A comparison of (6.8) with (6.2) shows that it is largely advantageous to bound directly the oscillations of ψ_n. This is what we do in this section by adapting the arguments of Section 5.

6.1 Coupling of the transition probabilities for blocks

For every integer n, we define a family of transition probability $\overline{P}_n$ on $(A^n)^2 \times \underline{A}^2$ by

$$\widetilde{P}_n(a_{0,n-1}, b_{0,n-1} \,|\, x, y) \;=\; \big[P_n(\cdot \,|\, x) \,\tilde{\times}\, P_n(.\,|\, y)\big](a_{0,n-1}; b_{0,n-1}) \;. \tag{6.9}$$

Let $(n_m)_{m\in\mathbb{N}}$ be an increasing sequence. For each double history x, y, we consider the coupling $\big((\overline{U}^{x,y}, \overline{V}^{x,y})\big)_{m\in\mathbb{Z}}$ of the chains for n_m-blocks with past x and y, defined by,

$$\begin{aligned}
&\mathbf{P}(\overline{U}^{x,y}_{0,n_m} = a_{0,n_m} \,,\, \overline{V}^{x,y}_{0,n_m} = b_{0,n_m}) \\
&\quad = \prod_{m=1}^{M} \overline{P}_{n_{m+1}-n_m}(a_{n_m,n_{m+1}} \,,\, b_{n_m,n_{m+1}} \,|\, a_{n_m} \cdots a_0 x \,,\, b_{n_m} \cdots b_0 y) \;.
\end{aligned} \tag{6.10}$$

6.2 The process of last block-differences

We set

$$\gamma^{(n)}_k \;=\; 1 - \inf\left\{ \frac{P_n(a_{0,n-1} \,|\, x)}{P_n(a_{0,n-1} \,|\, y)} \;:\; x \stackrel{k}{=} y \,,\; a_1, \ldots, a_{n-1} \in A \right\} \;. \tag{6.11}$$

From (4.11) we see that, for $x \stackrel{k}{=} y$, the weight of the diagonal of each coupling $\overline{P}_n$ satisfies

$$\Delta\big(\widetilde{P}_n(\cdot,\cdot \,|\, x, y)\big) \;\geq\; \inf_{a_0,\ldots,a_{n-1}\in A} \frac{P(a_{0,n-1} \,|\, x)}{P(a_{0,n-1} \,|\, y)} \;\geq\; 1 - \gamma^{(n)}_k \;. \tag{6.12}$$

If we denote

$$\overline{\Delta}^{x,y}_{m,m+q} \;:=\; \left\{ \overline{U}^{x,y}_j = \overline{V}^{x,y}_j \,,\, n_m \leq j \leq n_{m+q} \right\} \,,$$

we deduce from (6.12) that

$$\mathbf{P}(\overline{\Delta}_{m+k+1} \,|\, \overline{\Delta}_{m,m+k}) \;\geq\; 1 - \gamma^{(n_{m+k+1}-n_{m+k})}_{n_{m+k}-n_m} \;. \tag{6.13}$$

We construct the process $(\overline{T}_n)_{n\in\mathbb{N}}$ with

$$\overline{T}_m^{x,y} \;=\; \inf\Big\{p \geq 0 \,:\, U_i^{x,y} \neq V_i^{x,y} \text{ for some } i, n_{m-p} \leq i \leq n_{m-p+1}\Big\} \;. \tag{6.14}$$

By (6.13), the conditional laws of this process satisfy,

$$\mathbf{P}(\overline{T}_{m+1} = k+1 \,|\, \overline{T}_m = k) \;\geq\; 1 - \gamma_{n_{m+k}-n_m}^{(n_{m+k+1}-n_{m+k})} \tag{6.15}$$

and

$$\mathbf{P}(\overline{T}_{m+1} = 0 \mid \overline{T}_m = k) \;\leq\; \gamma_{n_{m+k}-n_m}^{(n_{m+k+1}-n_{m+k})} \;. \tag{6.16}$$

6.3 The dominating Markov process

Let us choose the length of the blocks in such a way that the sequence $(n_m)_{m\in\mathbb{N}}$ is subadditive, i.e.

$$n_{m+k} - n_m \;\leq\; n_k \tag{6.17}$$

for $m, k \geq 0$, and that

$$\sup_{n\geq 0} \gamma_\ell^{(n)} \;<\; 1 \tag{6.18}$$

for all $\ell \geq 0$. These two properties together with (6.15)–(6.16) imply that, for all histories x and y,

$$\mathbf{P}(\overline{T}_{m+1}^{x,y} = k+1 \,|\, \overline{T}_m^{x,y} = k) \;\geq\; 1 - \overline{\gamma}_k \tag{6.19}$$

and

$$\mathbf{P}(\overline{T}_{m+1}^{x,y} = 0 \,|\, \overline{T}_m^{x,y} = k) \;\leq\; \overline{\gamma}_k. \tag{6.20}$$

with

$$\overline{\gamma}_k \;:=\; \sup_{n\geq 1} \gamma_{n_k}^{(n)} \;, \tag{6.21}$$

for $m \geq 1$.

We now define the "dominating" Markov chain $(S_n^{(\overline{\gamma})})_{n\in\mathbb{N}}$ as in (2.8)–(2.9). Lemma 1 yields

$$\mathbf{P}(\overline{T}_m^{x,y} = 0) \;\leq\; \mathbf{P}(\overline{S}_m = 0) \;\leq\; \overline{\gamma}_m^* \;. \tag{6.22}$$

Hence, if $n_m \leq n \leq n_{m+1}$,

$$\mathbf{P}(\overline{U}_n^{x,y} \neq \overline{V}_n^{x,y}) \;\leq\; \mathbf{P}(T_m^{x,y} = 0) \;\leq\; \overline{\gamma}_m^* \;. \tag{6.23}$$

6.4 Decay of correlations

We can now mimick the proof of Theorem 5 in terms of barred objects.

As $(\mathrm{var}_m(\phi))_{m\in\mathbb{N}}$ is summable, there exists a subadditive sequence $(n_m)_{m\in\mathbb{N}}$ such that the sequence α_m of the tails

$$\alpha_m \;=\; \sum_{k\geq n_m} \mathrm{var}_k(\phi) \tag{6.24}$$

is summable:

$$\sum_{m\geq 0} \alpha_m < +\infty \;. \tag{6.25}$$

The transitions for blocks of size n satisfy

$$\frac{P_n(a_{0,n-1}\,|\,x)}{P_n(a_{0,n-1}\,|\,y)} \;\geq\; e^{-\mathrm{var}_k(\psi_n)} \tag{6.26}$$

if $x \stackrel{k}{=} y$. But from (6.8), (6.7) and (6.2) we have

$$\begin{aligned} \mathrm{var}_k(\psi_n) &\leq \left(\sum_{m=k}^{k+n} + \sum_{m\geq k+n} + \sum_{m\geq k}\right)\mathrm{var}_m(\phi) \\ &\leq 3\sum_{m\geq k}\mathrm{var}_m(\phi)\;. \end{aligned} \tag{6.27}$$

Hence we can choose in (6.21)

$$\overline{\gamma}_k \;\leq\; 1 - e^{-3\alpha_k}\;, \tag{6.28}$$

a choice for which

$$\sum_{k\geq 1}\overline{\gamma}_k < +\infty\;. \tag{6.29}$$

To prove the theorem, we now proceed as in (5.1) and (5.4)–(5.10) but replacing tildes by bars and putting bars over the processes (T_n) and $(S_n^{(\gamma)})$. We just point out that, due to the subadditivity of n_m,

$$\mathrm{var}_{(n_{m+k}-n_m)}(\phi) \;\leq\; \mathrm{var}_{n_k}(\phi)$$

uniformly in m. $\square$

A Returns to the origin of the dominating Markov chain

In this appendix we collect a few results concerning the probability of return to the origin of the Markov chain $(S_n^{(\gamma)})_{n\in\mathbb{N}}$ defined via (2.9). (In the sequel we omit the superscript "(γ)" for simplicity.)

Proposition 2 *Let $(\gamma_n)_{n\in\mathbb{N}}$ be a real-valued sequence decreasing to 0 as $n\to+\infty$.*

(i) *If $\sum_{m\geq 1}\prod_{k=0}^{m}(1-\gamma_k) = +\infty$, then $\mathbf{P}(S_n=0)\to 0$.*

(ii) *If $\sum_{m\geq 1}\gamma_k < +\infty$, then $\sum_{n\geq 0}\mathbf{P}(S_n=0) < +\infty$.*

(iii) *If (γ_m) decreases exponentially, then so does $\mathbf{P}(S_n=0)$.*

(iv) *If (γ_m) decreases polynomially, then $\mathbf{P}(S_n=0) = O(\gamma_n)$.*

Sketch of the proof

Statement (i) follows from the well known fact that the Markov chain $(S_n)_{n\in\mathbb{N}}$ is positive recurent if and only if,

$$\sum_{m\geq 1}\prod_{k=0}^{m}(1-\gamma_k) < +\infty.$$

To prove parts (ii) and (iii) we introduce the series

$$F(s) = \sum_{n=1}^{+\infty} \mathbf{P}(\tau = n)\, s^n \ , \tag{A.1}$$

and

$$G(s) \ = \ \sum_{n=0}^{+\infty} \mathbf{P}(S_n = 0)\, s^n \tag{A.2}$$

where the random variable τ is the time of first return to zero, defined in (5.8). The probabilities $\mathbf{P}(\tau = n)$ were computed in (5.11) above. The relation (5.7) implies that these series are related in the form

$$G(s) = \frac{1}{1 - F(s)} \ , \tag{A.3}$$

for all $s \geq 0$ such that $F(s) < 1$.

It is clear that the radius of convergence of F is at least 1. In fact,

$$F(1) \ = \ \mathbf{P}(\tau < +\infty) \ . \tag{A.4}$$

Moreover, if $\sum_{m\geq 1} \gamma_k < +\infty$, the radius of convergence of F is

$$\lim_{n\to\infty} [\gamma_n]^{-1/n} \ . \tag{A.5}$$

This is a consequence of the fact that $\mathbf{P}(\tau = n)/\gamma_{n-1} \to \mathbf{P}(\tau = +\infty) > 0$, as concluded from (5.11).

Statement (*ii*) of the proposition is a consequence of the fact that the radius of convergence of the series G is at least 1 if $\sum_{m\geq 1} \gamma_k < +\infty$. This follows from the relation (A.3) and the fact that the right-hand side of (A.4) is strictly less than one when the chain $(S_n^{(\gamma)})$ is transient.

To prove statement (iii) let us assume that $\gamma_m \leq C\gamma^m$ for some constants $C < +\infty$ and $0 < \gamma < 1$. By (A.5), the radius of convergence of F is $\gamma^{-1} > 1$ while, by (A.4), $F(1) < 1$. By continuity it follows that there exists $s_0 > 1$ such that $F(s_0) = 1$ and, hence, by (A.3), $G(s) < +\infty$ for all $s < s_0$. By definition of G, this implies that $\mathbf{P}(S_n = 0)$ decreases faster than ζ^n for any $\zeta \in (1, s_0^{-1})$.

Statement (iv) is a consequence of the following lemma.

Lemma 2 *If*

$$\alpha \ := \ \sup_i \overline{\lim}_{k\to\infty} \left[\frac{\mathbf{P}(\tau = i)}{\mathbf{P}(\tau = ki)}\right]^{1/k} \ < \ \frac{1}{\mathbf{P}(\tau < +\infty)} \ , \tag{A.6}$$

then

$$\mathbf{P}(S_n = 0) = O\left(\mathbf{P}(\tau = n)\right).$$

Proof We start with the following observation. If $i_1 + \cdots + i_k = n$, then $\max_{1\leq m\leq k} i_m.n/k$ and thus, for g is an increasing

$$g(n) \ \leq \ g\left(k\, i_{\max}\right) \ ,$$

where $i_{\max} = \max_{1\leq m\leq k} i_m$. If we apply this to $g(n) = 1/\mathbf{P}(\tau = n)$, which is increasing by (5.11), we obtain

$$1 \ \leq \ \frac{\mathbf{P}(\tau = n)}{\mathbf{P}(\tau = k\, i_{\max})} \ . \tag{A.7}$$

We now invoke the following a explicit relation between the coefficients of F and G.

$$\mathbf{P}(S_n = 0) = \sum_{k=1}^{n} \sum_{\substack{i_1, \ldots, i_k \geq 1 \\ i_1 + \cdots + i_k = n}} \prod_{m=1}^{k} \mathbf{P}(\tau = i_m) , \tag{A.8}$$

for $n \geq 1$. Multiplying and dividing each factor in the rightmost product by $\mathbf{P}(\tau < +\infty)$, this formula can be rewritten as

$$\mathbf{P}(S_n = 0) = \sum_{k=1}^{n} \mathbf{P}(\tau < +\infty)^k \sum_{\substack{i_1, \ldots, i_k \geq 1 \\ i_1 + \cdots + i_k = n}} \prod_{m=1}^{k} \mathbf{P}(\tau = i_m \,|\, \tau < +\infty). \tag{A.9}$$

Combining this with (A.7) we obtain

$$\mathbf{P}(S_n = 0) \;\leq\; \mathbf{P}(\tau = n) \sum_{k=1}^{n} \mathbf{P}(\tau < +\infty)^k \sum_{\substack{i_1, \ldots, i_k \geq 1 \\ i_1 + \cdots + i_k = n}} \prod_{m=1}^{k} \frac{\mathbf{P}(\tau = i_m \,|\, \tau < +\infty)}{\mathbf{P}(\tau = k\, i_{\max})} , \tag{A.10}$$

If we single out the factor $\mathbf{P}(\tau = i_{\max} \,|\, \tau < +\infty) = \mathbf{P}(\tau = i_{\max})/\mathbf{P}(\tau < +\infty)$ from the rightmost product of (A.10) and use the hypothesis (A.6) we get

$$\mathbf{P}(S_n = 0) \;\leq\; C\,\mathbf{P}(\tau = n) \sum_{k=1}^{n} \alpha^k\, \mathbf{P}(\tau < +\infty)^{k-1} \sum_{\substack{i_1, \ldots, i_k \geq 1 \\ i_1 + \cdots + i_k = n}} \prod_{\substack{1 \leq m \leq k \\ i_m \neq i_{\max}}} \mathbf{P}(\tau = i_m \,|\, \tau < +\infty) , \tag{A.11}$$

for some constant $C > 0$. To bound the last sum on the right-hand side we introduce a sequence of independent random variables $(\tau^{(i)})_{i \in \mathbb{N}}$ with common distribution

$$\mathbf{P}(\tau^{(i)} = j) \;=\; \mathbf{P}(\tau = j \,|\, \tau < +\infty) . \tag{A.12}$$

Then

$$\sum_{j=1}^{n-k+1} \mathbf{P}\Big(\sum_{i=1}^{k-1} \tau^{(i)} = n - j\Big) \;\leq\; 1 . \tag{A.13}$$

Hence, (A.11) implies

$$\mathbf{P}(S_n = 0) \;\leq\; C\,\alpha \sum_{k=1}^{\infty} \big[\alpha^{-1}\mathbf{P}(\tau < +\infty)\big]^{k-1}\; \mathbf{P}(\tau = n) \;\leq\; \text{const}\,\mathbf{P}(\tau = n) . \,\Box \tag{A.14}$$

We notice that, according to (5.11), $\gamma_n \sim \mathbf{P}(\tau = n)/\mathbf{P}(\tau = +\infty)$. Hence, a sufficient condition for (A.6) is a similar condition for the sequence (γ_n). Such a condition holds, for instance, if the later sequence decays polynomially. Statement *(iii)* of the proposition follows. □

Acknowledgements

It is a pleasure to thank Pablo Ferrari and Luis Renato Fontes for useful discussions. We also thank Pablo Ferrari for informing us about his work in progress with Alejandro Maass and Servet Marínez on a related regenerative representation of chains with complete connections.

References

[1] M BARNSLEY, S. DEMKO, J. ELTON and J. GERINOMO (1988). Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities. *Ann. Inst. Henri Poincaré, Prob. Statist.*, 24:367–394.

[2] H. BERBEE (1987). Chains with infinite connections: Uniqueness and Markov representation. *Probab. Th. Rel. Fields*, 76:243–253.

[3] R. BOWEN (1975). *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470. Springer.

[4] X. BRESSAUD, R. FERNANDEZ, and A. GALVES (1998). Speed of $\overline{d}$-convergence for Markov approximations of chains with complete connections. A coupling approach. To be published in *Stoch. Proc. and Appl.* A preliminary version can be retreived from `http://mpej.unige.ch/mp_arc/c/97/97-587.ps.gz`.

[5] K. BURZDY and W. KENDALL (1998). Efficient Markovian couplings: examples and counterexamples. Preprint. Can be retrieved from `http://www.math.washington.edu/~burdzy/Papers/list.html`

[6] W. DOEBLIN (1938). Exposé sur la théorie des chaines simples constantes de Markoff à un nombre fini d'états. *Rev. Math. Union Interbalkanique*, 2:77–105.

[7] W. DOEBLIN (1940). Remarques sur la théorie métrique des fractions continues. *Composition Math.*, 7:353–371.

[8] W. DOEBLIN and R. FORTET (1937). Sur les chaînes á liaisons complétes. *Bull. Soc. Math. France*, 65:132–148.

[9] T. E. HARRIS (1955). On chains of infinite order. *Pacific J. Math.*, 5:707–724.

[10] M. IOSIFESCU (1978). Recent advances in the metric theory of continued fractions. *Trans. 8th Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, 27–40. Academia, Prague.

[11] M. IOSIFESCU (1992). A coupling method in the theory of dependence with complete connections according to Doeblin. *Rev. Roum. Math. Pures et Appl.*, 37:59–65.

[12] M. IOSIFESCU and S. GRIGORESCU (1990). *Dependence with Complete Connections and its Applications*, Cambridge University Press, Cambridge.

[13] M. IOSIFESCU and R. THEODORESCU (1969). *Random Processes and Learning*, Springer-Verlag, Berlin.

[14] T. KAIJSER (1981). On a new contraction condition for random systems with conplete connections. *Rev. Roum. Math. Pures et Appl.*, 26:1075–1117.

[15] T. KAIJSER (1994). On a theorem of Karlin. *Acta Applicandae Mathematicae*, 34:51–69.

[16] S. KARLIN (1953). Some random walks arising in learning models. *J. Pacific Math.*, 3:725–756.

[17] A. KONDAH, V. MAUME and B. SCHMITT (1996). Vitesse de convergence vers l'état d'équilibre pour des dynamiques markoviennes non höldériennes. Technical Report 88, Université de Bourgogne.

[18] S. P. LALLEY (1986). Regeneration representation for one-dimensional Gibbs states. *Ann. Prob.*, 14:1262–1271.

[19] F. LEDRAPPIER (1974). Principe variationnel et systèmes dynamiques symboliques. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 30:185–202.

[20] T. LINDVALL (1991). W. Doeblin 1915–1940. *Ann. Prob.*, 19:929–934.

[21] T. LINDVALL (1992). *Lectures on the Coupling Method*, Wiley, New York.

[22] C. LIVERANI (1995). Decay of correlations. *Ann. of Math.*, 142(2):239–301.

[23] F. NORMAN (1972). *Markov Processe and Learning Models*, Academic Press, New York.

[24] O. ONICESCU and G. MIHOC (1935). Sur les chaines statistiques. *C. R. Acad. Sci. Paris*, 200:5121—512.

[25] O. ONICESCU and G. MIHOC (1935a). Sur les chaines de variables statistiques. *Bull. Sci., Math.*, 59:174–192.

[26] W. PARRY and M. POLLICOTT (1990). Zeta functions and the periodic structure of hyperbolic dynamics. *Asterisque* 187-188.

[27] M. POLLICOTT (1997). Rates of mixing for potentials of summable variation. Preprint.

[28] A. N. QUAS (1996). Non-ergodicity for c^1 expanding maps and g-measures. *Ergod. Th. Dynam. Sys.*, 16:531–543.

[29] D. RUELLE (1978). *Thermodynamic formalism.* Encyclopedia of Mathematics and its applications, volume 5. Addison-Wesley, Reading, Massachusetts.

[30] P. WALTERS (1975). Ruelle's operator theorem and g-measures. *Trans. Amer. Math. Soc.*, 214:375–387.

[31] P. WALTERS (1978). Invariant measures and equilibrium states for some mappings which expand distances. *Trans. Amer. Math. Soc.*, 236:121–153.

[32] L.-S. YOUNG (1997). Decay of correlations for certain quadratic maps. *Commun. Math. Phys.*, 146:123–138.

Xavier Bressaud and Antonio Galves
Instituto de Matemática e Estatística
Universidade de São Paulo
Caixa Postal 66281
05315-970 São Paulo, Brasil
e-mail: `bressaud, galves@ime.usp.br`

Roberto Fernández
Instituto de Estudos Avançados
Universidade de São Paulo
Av. Prof. Luciano Gualberto, Travessa J, 374 Térreo,
Cidade Universitária
05508-900 São Paulo, Brasil
e-mail: `rf@ime.usp.br`